Naval Aerospace Medical
Research Laboratory

Preliminary Validation of a Readiness-to-Fly Assessment Tool for Use in Naval Aviation

J. F. Chandler, R. D. Arnold, J. B. Phillips,
Renee A. Lojewski, & Dain S. Horning

NAMRL Report Number 10-22

Reviewed and Approved
08-06-2010

Rita G. Simmons, CDR MSC USN

This work was funded by work unit number 70803.

The study protocol was approved by the Naval Aerospace Medical Research Laboratory Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

EXECUTIVE SUMMARY

Fatigue is the most frequently cited physiological factor contributing to the occurrence of US Naval Aviation flight mishaps (Naval Safety Center, 2006). The Navy and other military services have invested significant resources in the development of means to manage and mitigate fatigue in operational settings. Foremost among these investments is the development of fatigue modeling/scheduling tools, the primary function of which is to inform mission scheduling to minimize fatigue and improve safety and operational effectiveness. Although *generalized* fatigue modeling tools, such as the Fatigue Avoidance Scheduling Tool (FAST), are increasingly used in military settings, currently there is no established tool available to assess an *individual* aviator's actual real-time level of fatigue or general physiological readiness. Recent evidence suggests large individual differences in fatigue resistance exist (Van Dongen, Caldwell, & Caldwell, 2006; Killgore, Grugle, Reichardt, Killgore, & Balkin, 2009), pointing to the need to supplement general models of fatigue with individualized fatigue measurement and modeling. Accordingly, the Naval Safety Center (NSC) has identified the need for a quickly-administered individualized fatigue assessment tool to determine a pilot or aircrew member's readiness-to-fly.

In response to this need, NAMRL was funded by the Bureau of Medicine and Surgery (BUMED) Medical Development Program to conduct validation research of several cognitive and physiological test instruments for their potential to serve as individualized fatigue detection tools. The instruments evaluated included Flight Fit, a brief (appx. 7 to 8 minute) computer-based cognitive test battery. Flight Fit is composed of tasks measuring cognitive abilities crucial for handling heavy mental work load and sensitive to the effects of fatigue (e.g., time-estimation, decision-making, short-term memory). The second primary instrument evaluated was PMI Fit 2000, which measures several oculometric characteristics putatively sensitive to the effects of fatigue, including, pupil diameter, pupil constriction amplitude and latency, and saccadic velocity. In addition to the two main instruments, the Psychomotor Vigilance Task (PVT), a gold standard in detecting fatigue; Synthetic Work for Windows (SynWin), a test of working memory and cognitive load; simulated flight performance with X-Plane 9, an ecologically valid, aviation-specific, measure of vigilance; and the Stanford Sleepiness Scale, a subjective assessment of sleepiness were evaluated For purposes of secondary analysis, performance was predicted using fatigue and performance modeling software, the Fatigue Avoidance Scheduling Tool (FAST). Subjects' baseline sleep/wake data were collected via actigraphy and entered into the FAST models.

Fifteen study participants were observed over a three day period. During days one and two, baseline test performance data were collected, in addition to actigraphic data on participants' sleep/wake patterns. Day three consisted of a 25 hour period of continual wakefulness (0300 hours to 0400 hours), during which test and performance data were collected at three hour intervals. It was hypothesized that over the course of 25 hours of continual wakefulness, participants would exhibit decrements on cognitive (Flight Fit) and physiological (PMI FIT 2000) measures. Additionally, it was hypothesized that secondary validity indices would demonstrate concomitant performance decrements due to sleep loss, evidenced through performance on the PVT, SynWin, and X-Plane flight simulator, and through reports of subjective sleepiness on the SSS. Although it was anticipated that group performance decrements would be predicted by FAST modeling, it was hypothesized that some measures of fatigue would exhibit significant individual differences, and that the addition of these measures would incrementally improve the prediction of fatigued task performance over FAST alone.

Analyses and results are discussed in detail in three Stages. Stage 1 establishes the relation of significant measures to decrement across time spent without sleep, and therefore fatigue. Stage 2 further explores significant Stage 1 relations as predictors of fatigue-related performance (PVT lapses) at group and individual levels. Stage 3 uses results from Stages 1 and 2 to inform the construction of optimal group scoring algorithms to predict fatigue-related performance. Aspects of Flight Fit and PMI Fit 2000 showed significant predictive ability across all three Stages of analysis, with individual variability playing a significant role when examined in Stage 2. The findings suggest that basic cognitive *and* physiologic tasks can successfully measure fatigue, and that both are necessary for optimal measurement. Further, scores on subsets of these same tasks can differentiate an individual's personal level of fatigue susceptibility above and beyond the current industry standard tool. Finally, combining the individual diagnostic power of Flight Fit and PMI Fit 2000 with established group measures such as FAST elicits greater

predictive ability of fatigued performance than either approach alone. These results are based on analysis of raw data from Flight Fit and PMI Fit 2000; normative scores provided by the manufacturers' algorithms did not yield significant results. Therefore, the current algorithms used to score Flight Fit and PMI Fit 2000 must be adjusted to reflect use in a Naval Aviation population. With that adjustment, both Flight Fit and PMI Fit 2000 show promise as valid real-time readiness-to-fly assessment tools in Naval Aviation squadrons. Follow-on studies to address scoring adjustment, as well as validation in a wider array of fatigue conditions (i.e., chronic, cumulative sleep debt) are discussed.

INTRODUCTION

The negative impact of fatigue is well known. Fatigue due to sleep loss causes slowed physiological and cognitive reaction time, memory problems, and increased mistakes during even routine decision making (Caldwell et al, 2009). Pilots and aircrew routinely report feeling fatigued in the cockpit (e.g., Belland & Bissell, 1994), and on an objective level, pronounced neurological and physiological decrements have been associated with both chronic (Van Dongen, Rogers, & Dinges, 2003) and acute (Caldwell, 2005) sleep loss. The Navy has documented fatigue issues in relation to aviation for over 50 years (see Graybiel, Brown, & Crispell,1943; Graybiel, Horwitz, & Gates, 1944). Today, fatigue is the most frequently cited physiological factor contributing to the occurrence of US Naval Aviation flight mishaps (Naval Safety Center, 2006), costing hundreds of millions of dollars in lost equipment and the incalculable cost of lost human life.

Accordingly, the Navy and other military services have invested significant resources in the development of means to manage and mitigate fatigue in operational settings. Mitigation techniques have largely focused on pharmacologic countermeasures, for example, caffeine and dextroamphetamine. Pharmacologic interventions continue to improve, increasing efficacy while decreasing adverse side effects. An excellent example of such an intervention is the drug modafanil and its extended action formulation, armodafanil (Phillips, Arnold, Strompolis, & Simmons, 2009). Modafanil and its variant have demonstrated many of the benefits of traditional stimulants already used by military communities without severely affecting normal sleep patterns or appetite. Modafanil also appears to have a lower potential for abuse than currently used stimulants (Lyons & French, 1991; Myrick, Malcom, Taylor, & LaRow, 2004). Even with the advancement of mitigation techniques, *prevention* of the fatigue state remains ideal. Efforts at prevention are centered on fatigue management through the use of predictive modeling and scheduling tools. These include duty hours, rotations and flight times used to inform mission scheduling to minimize fatigue and improve safety and operational effectiveness. In a position paper recently adopted by the Aerospace Medical Association, Caldwell and colleagues (2009) outline two major types of fatigue management technologies: 1) on-line real time assessment, and 2) off-line fatigue prediction algorithms.

On-line, real-time assessment of fatigue focuses on continuous tracking of physiologic markers sensitive to fatigue to calculate when an individual falls below an acceptable level of alertness. For instance, the Percentage of Eye Closure (PERCLOS) metric assesses the frequency and duration of slow eye blinks, a behavior strongly linked to drowsiness and sleep loss-related fatigue (Dinges, Mallis Maislin, & Powell, 1998). The PERCLOS metric is employed in several currently available commercial fatigue detectors that all operate similarly. Once an undesirable level is reached in the PERCLOS metric, the operator is notified that the subject has reached an unsafe fatigue level, usually by an alarm. The alarm is intended to reorient the individual long enough to take effective countermeasures. Other physiologic indicators, such as eye gaze, head position, actigraphy, and EEG have been employed in real time monitoring systems to good effect (Caldwell et al, 2009). However, the difficulty of positioning and maintaining these systems in high tempo operational environments makes their transition to military application problematic. For example, PERCLOS and other oculometrics rely on relatively stable head positioning for accuracy, a functionally impossible condition in the cockpit. Further, detection of fatigue at its onset is not operationally ideal; by the time deficits are detectable, performance is already compromised. This consideration has driven the development of fatigue prediction algorithms, so that performance deficits may be anticipated before they occur.

The use of off-line fatigue prediction algorithms is well illustrated by the Fatigue Avoidance Scheduling Tool (FAST), a program designed to measure, estimate, and manage performance changes induced by sleep restriction or deprivation and time of day. The performance predictions are based on the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE™) Model, extensive field data, and sleep deprivation studies (Hursh et al., 2004). The output for FAST includes a prediction of performance effectiveness represented as a relative departure from baseline functioning across the course of a day. For ease of interpretation, the performance scores given by FAST can be equated to blood alcohol level (BAL), a metric with well-known cognitive and physiological performance correlates. The FAST software can be used with actigraphy or data can be entered from a self-report of the individual's sleep/wake cycle. Performance prediction by FAST includes a few key assumptions, most notably that

all individuals have highly similar circadian rhythms and fatigue responses. While these assumptions are based on group normative data, recent evidence suggests large individual differences in fatigue resistance exist (Van Dongen, Baynard, Maislin & Dinges, 2004), and that these differences may be connected to aspects of basic cognitive functioning (Killgore et al., 2009).

Inter-individual variations in fatigue response highlight the need to supplement general models of fatigue, such as the SAFTE™ Model, with individualized fatigue measurement and modeling using a combination of cognitive and physiological factors. Currently there is not an established tool available to assess an *individual* aviator's actual real-time level of fatigue or general physiological readiness in this capacity. Accordingly, the Naval Safety Center (NSC) has identified the need for a quickly-administered individualized fatigue assessment tool to determine a pilot or aircrew member's readiness-to-fly. The current report documents testing of two potential instruments to fill that need, the Flight Fit cognitive fatigue assessment and the PMI FIT 2000 physiological fatigue assessment tools.

METHOD

Subjects

Fifteen active duty military personnel from the Naval Aviation Preflight Indoctrination (API) program volunteered as test subjects. The study protocol was approved by the Naval Aerospace Medical Research Laboratory Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects. Descriptive statistics for the subjects are presented in Table 1.

No specific groups were excluded. However, certain factors identified via a medical history form (Appendix B), served to exclude individual participants, due to their potential confounding effects. These included excessive alcohol use within the previous 48 hours (>3 drinks), greater than 400mg of routine daily caffeine consumption, habitual use of tobacco products within the previous six months, and history of significant medical, neurological, psychiatric, or sleep-related problems (Killgore, et al., 2009).

Table 1. Descriptive Statistics

	Age (years)		Height (in)		Weight (lbs)	
	Mean	SD	Mean	SD	Mean	SD
Male (n = 13)	24.7	2.1	71.2	3.3	186.6	20.0
Female (n = 2)	21.5	0.7	66.5	3.5	142.5	17.7
Total	24.3	2.3	70.5	3.6	180.7	24.6

Ethnicity	White	Black	Asian American	Hispanic/Latino(a)	Other
	11	2	0	2	0

Fatigue Assessments

Flight Fit*.* The Flight Fit cognitive test battery is an abbreviated (7 to 8 minute) version of the CogniFit assessment battery (full version is approximately 30 minutes) (Cognifit Inc., Yoqneam Ilit, Israel). The test measures cognitive performance on various components of mental work load sensitive to the effects of fatigue. Specifically, Flight Fit (FF) measures raw reaction time (FF_rawRT), visual scanning reaction time (FF_vsRT), visual scanning accuracy (FF_vsACC), divided attention reaction time (FF_daRT), divided attention accuracy (FF_daACC), shifting reaction time (FF_SRT), attention shifting accuracy (FF_shiftACC), focus reaction time in

the presence of distracters (FF_fdRT), and short-term memory (FF_STM) (see Appendix A for a complete listing of variable abbreviations).

PMI Fit 2000. The PMI FIT 2000 (PMI Inc., Rockville, MD) uses eye-tracking and pupillometry to identify impaired physiological states due to fatigue and other factors, such as alcohol or drug use. The test requires less than one minute to complete. The system employs an algorithm that compares an individual's established baseline to present state on 4 variables (i.e., pupil diameter, pupil constriction amplitude, pupil constriction latency & saccadic velocity). The baseline is established by the average of 10 trials taken during non-impaired conditions. After the baseline trials, each subsequent trial provides the user with scores on the four test components plus a composite score, the FIT Index. The PMI FIT 2000 has been used in multiple fatigue and impairment studies in other contexts, such as motor vehicle operation, and has been demonstrated to be both reliable and valid (e.g., Russo et al., 1999).

Psychomotor Vigilance Task. The PVT-192 (Ambulatory Monitoring Inc., Ardsley, New York) is a brief vigilance and attention task, and is considered the gold standard instrument for assessment of the effects of fatigue (Balkin et al., 2004). During each 10 minute trial, subjects are required to attend closely to a stimulus window and respond by pressing a response button. Subjects are instructed to respond as quickly as possible. PVT scores of interest include mean reciprocal reaction time of the slowest 10% of responses (Mean S RRT), and lapses (responses to stimulus presentations taking longer than 500 ms).

SynWin. SynWin is a computer-based test module that simulates a work environment by presenting up to four tasks on the screen simultaneously. These tasks include versions of the Sternberg Memory Task, mathematical calculation, gauge monitoring, and auditory vigilance. Each 10 minute trial is scored individually, and combines the participant's performance on all administered tasks into a single proficiency score.

Flight Simulation (X-Plane 9). Simulated flight performance was measured using the X-Plane 9 (Laminar Research) flight simulator. Because fatigue impairs basic attentional processes, simple tasks which are subject to more reliable measurement were the focus of simulated flight performance. Specifically, subjects were given a simple flight profile, with instructions to fly "straight and level" at a specified altitude, airspeed and heading (i.e., 2000 ft, 140 knots, due North). Deviations from these specified flight parameters were assessed.

Stanford Sleepiness Scale. Subjective sleepiness was assessed with the Stanford Sleepiness Scale (SSS), (Hoddes, Dement & Zarcone, 1972). The available scores for the SSS range from 1 ("Feeling active, vital, alert, or wide awake) to 7 ("No longer fighting sleep, sleep onset soon; having dream like thoughts"). There is also a means to denote if the subject is "Asleep", with the score of "X". The SSS is a widely used, easy-to-administer paper-and-pencil measure and has demonstrated excellent sensitivity to the effects of fatigue (Balkin et al., 2004).

Fatigue Avoidance Scheduling Tool. The Fatigue Avoidance Scheduling Tool (FAST; Nova Scientific Corporation, Fairborn, OH) is software designed to measure, estimate and manage performance changes induced by sleep restriction or deprivation and time of day. The primary use of FAST is to optimize the operational management of aviation crews and to design work schedules and mission-critical events in a manner that will reduce fatigue and fatigue induced errors. The performance predictions are based on the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE™) Model, numerous laboratory collaborations, field data collection, and sleep deprivation studies (Hursh et al., 2004). The output for FAST includes a prediction of performance effectiveness, which can also be used to extrapolate a blood alcohol level (BAL). The FAST software can be used with actigraphy or data can be entered from a self-report of the individual's sleep/wake cycle.

Design

The experiment employed a repeated measures design to investigate the effects of sleep deprivation on physiological state and task performance over time. The experiment consisted of two phases, (1) the Practice Phase and (2) the Experimental/Sleep Deprivation Phase.

Procedures

Practice Phase. Up to four (4) volunteers were recruited during each week of the study. After receipt of participants' informed consent, the Practice Phase of the experiment began. This phase was executed Monday and Tuesday morning and required approximately 90 minutes of participation each day. Practice Phase data was used for each of the measures to establish performance asymptote and to mitigate practice effects during the Experimental/Sleep Deprivation phase. Each day participants completed: 5 trials of the PMI FIT 2000, 2 trials of Flight Fit, 2 trials of the PVT, three 10-minute trials of SynWin, one 15-minute trial of the X-Plane Flight Simulator and the SSS. Prior to departing Monday morning, each subject was outfitted with a Motionlogger Microsleep Watch (Ambulatory Monitoring, Inc., Ardsley, New York), which was used to monitor sleep and wake periods while not under observation.

Experimental/Sleep Deprivation Phase. Upon completion of Tuesday morning Practice Phase, subjects were released with instructions to return at 0530 Wednesday morning. Subjects were instructed to sleep according to their normal schedules, and to awaken at 0300 Wednesday, remaining awake until the 0530 report time. Compliance was gauged by actigraphy. Subjects were also re-familiarized with the protocol for the sleep deprivation phase of the study. Beginning at 0600 subjects were assessed on Flight Fit, PMI FIT 2000, PVT, SynWin, SSS and the flight simulator task once every three (3) hours, as follows: 1 trial of PMI FIT 2000, 1 trial of Flight Fit, 1 trial of PVT, 1 administration of SSS, 1 trial of the simulated flight profile and 1 trial of SynWin. Trials began at 0600, 0900, 1200, 1500, 1800, 2100, 0000 (Thursday), and 0300. Upon completion of the final trial, subjects were debriefed and driven to the Bachelor Officers' Quarters (BOQ) with instructions to obtain adequate sleep prior to check out.

ANALYSES AND RESULTS

Overview

Three stages of data analysis were conducted in order to examine group and individual patterns of fatigue-related performance decrements. In Stage 1, a series of Repeated Measures ANOVAs was conducted for each criterion and predictor variable over the 8 Experimental Phase trials to determine which variables exhibited change across time. Significant change in predictor variables across time established their sensitivity to fatigue on a group level. Displaying change across time for criterion variables, such as PVT Lapses, is necessary in order to establish those variables as fatigue-related, and therefore appropriate as outcome measures for Stage 2 predictive models. In Stage 2, a series of Hierarchical Linear Models (HLMs) was conducted to predict performance decrements associated with fatigue, and to simultaneously examine any individual differences that were not evident at the group level analyses. Bivariate and multiple predictor models were examined. In Stage 3 we constructed several multiple predictor General Linear Models (GLMs) from significant Stage 2 predictor variables to formulate optimum group-based scoring algorithms for fatigue-related performance decrements using Flight Fit and PMI Fit 2000 components.

Stage 1

Stage 1 analyses were performed using SPSS version 16.0 for Windows (SPSS Inc., Chicago, IL). A series of Repeated Measures ANOVAs was conducted for each dependent variable over the 8 Experimental Phase trials. The 0600 trial of the Experimental/Sleep Deprivation Phase was established as baseline performance. A value of $p \leq 0.05$ was considered statistically significant. The following section describes each measure, and the variables assessed.

Predictor Variables

Flight Fit. There were 9 sub-scores for each trial of Flight Fit: FF_rawRT, FF_vsRT, FF_vsACC, FF_STM, FF_daACC, FF_SRT, FF_shiftACC, FF_fdRT, and FF_daRT. Results indicate that four Flight Fit sub-components detected significant fatigue effects across trials, including, FF_rawRT, FF_STM, FF_daACC and, FF_shiftACC. Post-hoc analyses revealed significant decreases in performance on these four measures indicative of fatigue effects, with the most dramatic detriments occurring during the last two assessment periods (0000 and 0300 hours). ANOVA results are presented in Table 2, and mean performance scores over assessment times for each significant sub-score are presented in Figures 1 – 4. Significant sub-components were retained in Stage 2 analyses to be evaluated as predictor variables.

Table 2. ANOVA results for Flight Fit Sub-Scores

	F	df	*p*	η_p^2
FF_rawRT	2.86	(7, 98)	.009	.17
FF_fdRT †	.82	(4.60, 64.52)	.533	.06
FF_vsACC †	1.64	(4.36, 60.97)	.172	.11
FF_vsRT	1.71	(7, 98)	.114	.11
FF_STM	2.45	(7, 98)	.023	.15
FF_daACC	2.41	(7, 98)	.026	.15
FF_daRT †	1.69	(3.60, 50.34)	.172	.11
FF_shiftACC	3.49	(7, 98)	.002	.20
FF_SRT	1.72	(7, 98)	.113	.11

† Geisser-Greenhouse correction used due to violation of sphericity

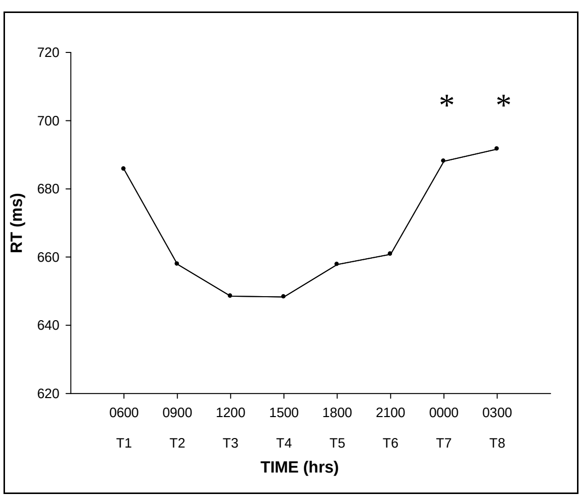

Figure 1. Mean Flight Fit Raw Reaction Time (FF_rawRT) scores in milliseconds at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T4 through T8; T5 and T1, T7, and T8; T6 and T1, T7, and T8; T7 and T1, T3, T4, T5, and T6; T8 and T1, T3, T4, T5, and T6. The most operationally significant differences, between T6 and T7 – T8, are noted (*).

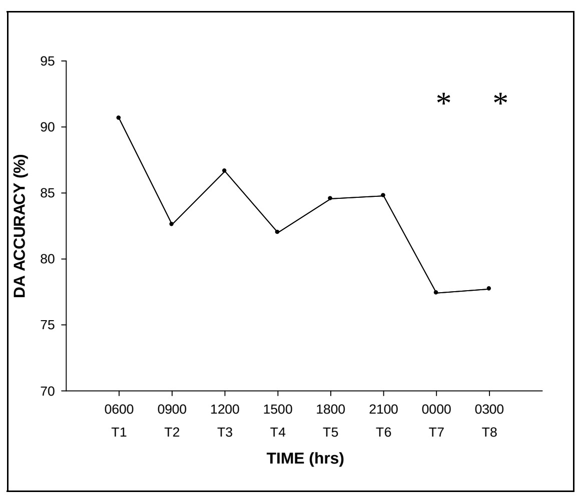

Figure 2. Mean Flight Fit Divided Attention Accuracy (FF_daACC) scores in percent correct at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T4 – T8; T5 and T1, T7, and T8; T6 and T1, T7, and T8; T7 and T1, T5, and T6; T8 and T1, T5, and T6. As with FF-rawRT, the most operationally significant differences, between T6 and T7 – T8, are noted (*).

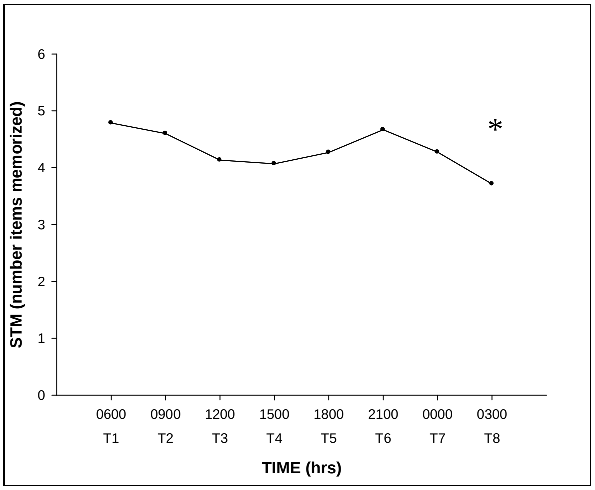

Figure 3. Mean Flight Fit Short Term Memory performance (FF_STM) in number of successfully memorized items at each test trial across time. Post-hoc analyses revealed significant differences between T8 and T1, T2, and T6, indicating that FF_STM performance was relatively stable until a significant decline at the 24 mark of continual wakefulness (*).

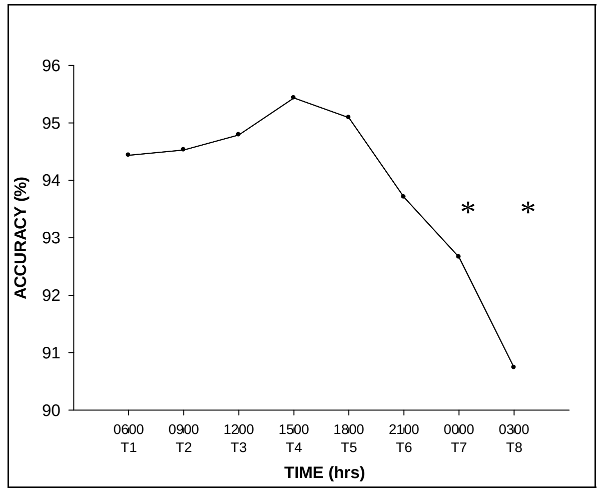

Figure 4. Mean Flight Fit Shifting Accuracy (FF_shiftACC) scores in percent correct at each test trial across time. Post-hoc analyses revealed significant differences between T4 and T7, T8; T8 and T2, T3, T4, T5, and T6, indicating that FF_shiftACC performance declined significantly and steadily from the T4 time slot until the final test trial (*).

PMI Fit 2000. There are four components of the FIT Index: pupil diameter, pupil constriction amplitude, pupil constriction latency and saccadic velocity. Results are displayed in Table 3 and Figures 5-9. Although the FIT Index failed to detect fatigue effects across the experimental time points, one subcomponent of the FIT Index, saccadic velocity, appears especially sensitive to the effects of fatigue (Figure 9). While significant effects were present for amplitude as well, examination of post-hoc tests revealed patterns that do not suggest that effects were associated with fatigue (Figure 6). As a result, only saccadic velocity was retained as a predictor variable in Stage 2 analyses.

Table 3. ANOVA results for PMI 2000

	F	df	p	η_p^2
Diameter†	.613	(3.79, 53.10)	.647	.042
Amplitude	4.93	(7, 98)	.000*	.260
Latency	1.41	(7, 98)	.211	.091
Saccadic Velocity†	8.88	(3.24, 45.34)	.000*	.388
FIT Index †	.774	(2.70, 37.76)	.503	.052

† Geisser-Greenhouse correction used due to violation of sphericity
* Mean difference significant at the .05 level

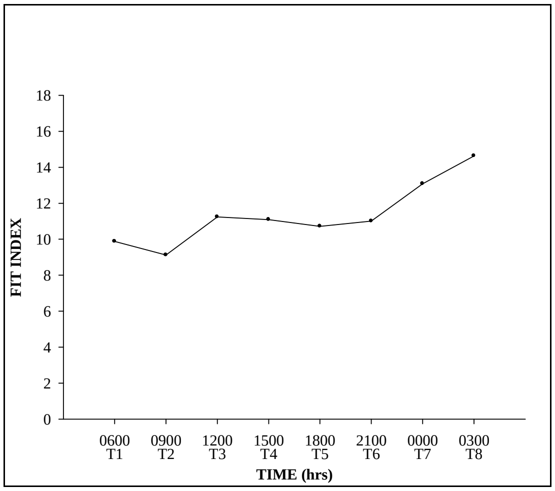

Figure 5. Mean PMI Fit Index Scores at each test trial across time. The omnibus test of the effect was not significant (Table 3); therefore, post-hoc analyses were not conducted.

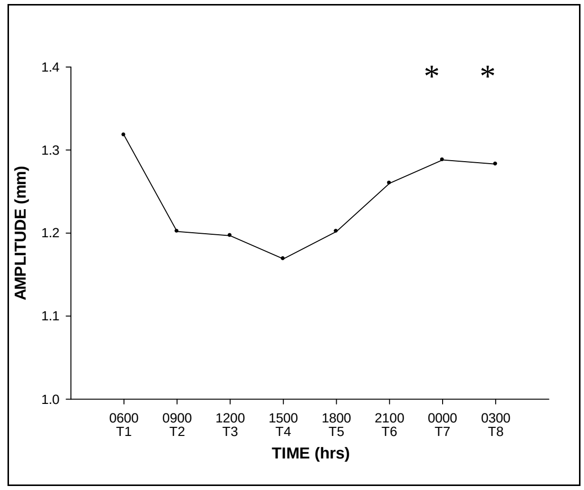

Figure 6. Mean PMI Pupil Constriction Amplitude in millimeters at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T2 – T6; T2 – T5 and T1, T7, and T8; T6 and T1; T7 and T2, T3, T4, and T5; T8 and T2, T4, and T5. The most operationally significant differences, between T5 and T7 – T8, are noted (*).

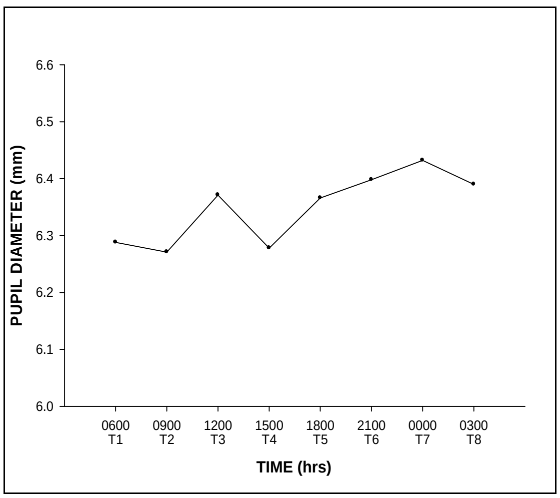

Figure 7. Mean PMI Pupil Diameter in millimeters at each test trial across time. The omnibus test of the effect was not significant (Table 3); therefore, post-hoc analyses were not conducted.

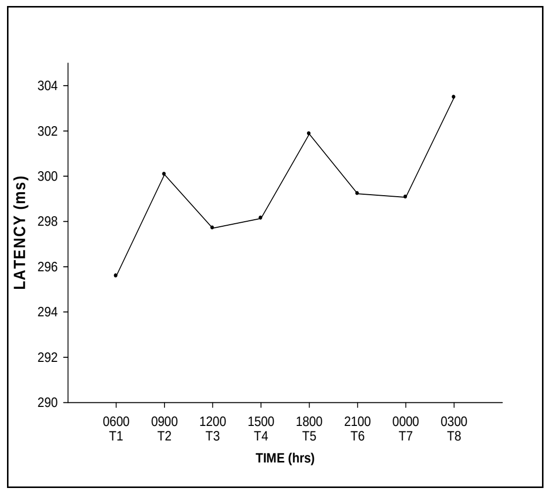

Figure 8. Mean PMI Pupil Constriction Latency in milliseconds at each test trial across time. The omnibus test of the effect was not significant (Table 3); therefore, post-hoc analyses were not conducted.

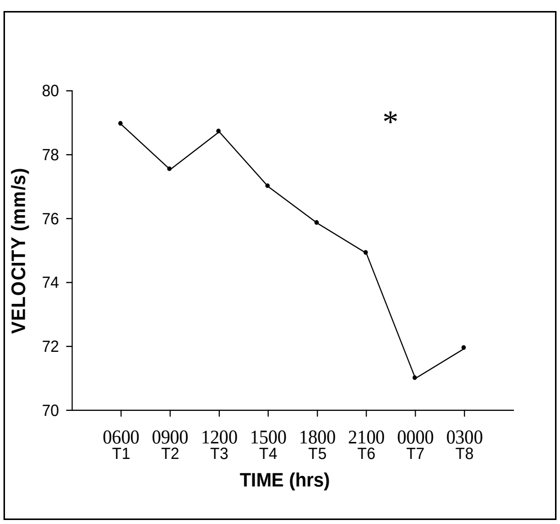

Figure 9. Mean PMI Saccadic Velocity (PMI_SV) in millimeters per second at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T7, T8; T2 and T6, T8; T3 and T6 – T8; T4 and T7, T8; T5 and T7, T8; T6 and T2, T3, T7, and T8; T7 and T1 – T6; T8 and T1 – T7, indicating that PMI_SV speed dropped significantly and steadily from the T3 trial until the T7 trial. The most operationally significant drop, from T6 to T7, is noted (*).

Criterion Variables

Psychomotor Vigilance Task. There were two variables of interest from the PVT: Mean S RRT and the number of lapses per trial. Results indicate significant fatigue effects for lapses and Mean S RRT, see Table 4 and Figures 10 and 11 for details. As lapses are both a fatigue literature gold standard and an operationally relevant vigilance analogue, it was included as the primary criterion variable in Stage 2 and 3 analyses.

Table 4. ANOVA results for PVT

	F	df	*p*	η_p^2
PVT Lapses†	6.88	(1.45, 20.28)	.009*	.329
Mean S RRT	9.36	(7, 98)	.000*	.401

† Geisser-Greenhouse correction used due to violation of sphericity
* Mean difference significant at the .05 level

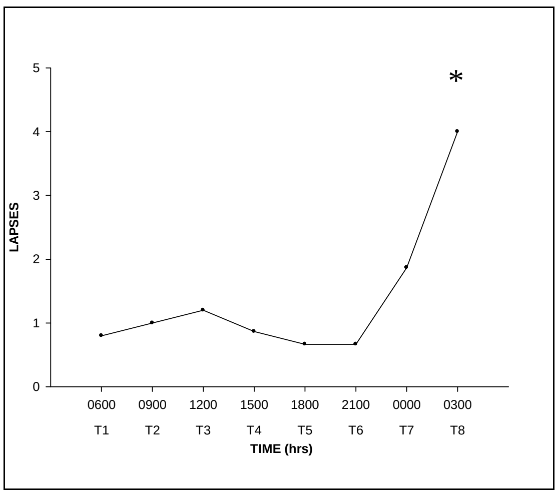

Figure 10. Mean PVT Lapses at each test trial across time. Post-hoc analyses revealed significant differences between T8 and all other trials, indicating a distinct point at which group vigilance began to fail (*).

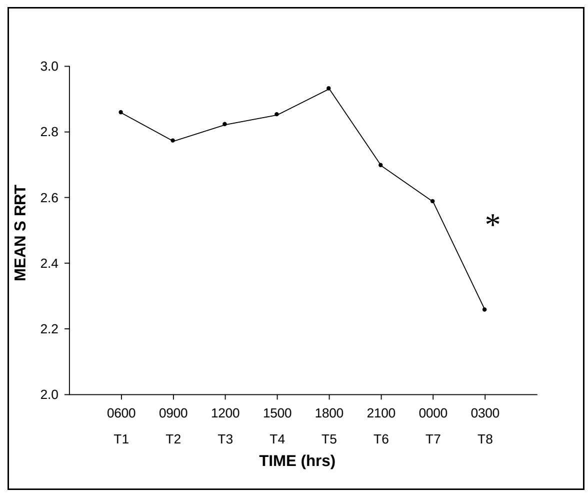

Figure 11. Mean reciprocal reaction time of the slowest 10% of responses (Mean S RRT) for the PVT at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T8; T2 and T8; T3 and T7, T8; T4 and T7, T8; T5 and T6, T7, and T8; T6 and T5, T8; T7 and T3, T4, T5, and T8; T8 and T1 – T7. This pattern is extremely similar to PVT lapses, with the final trial significantly slower than all other trials (*).

SynWin. Results revealing significant differences across time for the composite scores are displayed in Table 5 and Figure 12. Although significant effects were found for assessment time, these effects appear to be associated with dramatic performance variation from assessment to assessment as opposed to effects that can be clearly explained by fatigue (see figure 12). Though a marked performance decrement does appear from 2100 to 0300, the lack of consistency across time makes SynWin unsuitable as an outcome measure for further analyses.

Table 5. ANOVA results for SynWin composite score

F	df	*p*	η_p^2
4.70	(7,98)	.000*	.251

* Mean difference significant at the .05 level

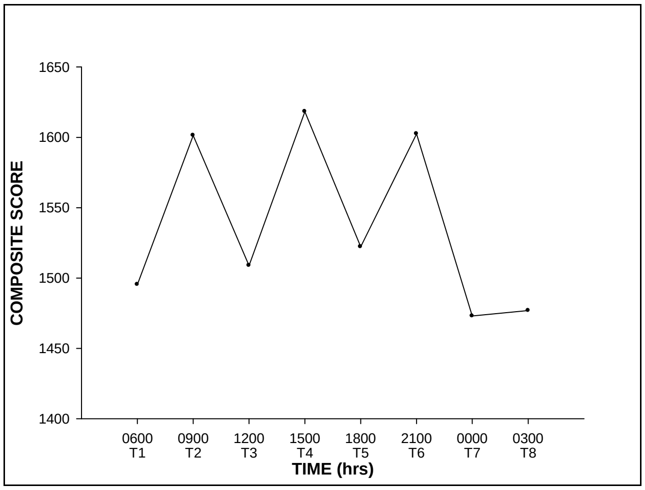

Figure 12. Mean SynWin Composite Score at each test trial across time. Post-hoc analyses revealed significant differences between T1 and T2, T4, and T6; T2 and T1, T7, and T8; T3 and T4, T6; T4 and T1, T3, T5, T7, and T8; T5 and T4, T6; T6 and T1, T3, T5, T7, and T8; T7 and T2, T4, and T6; T8 and T2, T4, and T6. The pattern of differences, though statistically significant, was not operationally interpretable in this context due to lack of consistency across time.

***Flight Simulation (X-Plane 9)*.**

Calculation of Total Lapse Time. Deviations from the specified flight parameter goals for heading (due North), airspeed (140 kts), and elevation (2000 ft) were calculated separately. Lapse times were calculated for each parameter as the number of seconds during a simulator trial that subjects deviated from the flight goal by greater than one standard deviation (determined at baseline). Total lapse time was the sum of lapse times for each parameter. The analysis revealed dramatic and significant effects of assessment time on total lapse time suggesting that total laps time is sensitive to fatigue effects. Results are displayed in Table 6 and Figure 13. These initial results suggest that it is possible to construct an ecologically valid measure of vigilance using low cost, commercially available flight simulation. Though promising, these are preliminary results only; further validation across time and varying situations is needed before Flight Simulator lapses can be used as an outcome measure with the same confidence as PVT lapses.

Table 6. ANOVA results for Flight Simulator Total Lapse Time†

F	df	*p*	η_p^2
2.53	(7, 98)	.02	.15

† Geisser-Greenhouse correction used due to violation of sphericity

***Stanford Sleepiness Scale (SSS)*.** Results for the SSS scores show that there was a significant main effect of assessment time. Post hoc comparisons showed significant differences between levels, the most revealing between Trials 6, 7, and 8 and all other Trials (see Table 7 and Figure 14), with individuals reporting greater sleepiness linearly across time. However, any self-reported subjective state has significant drawbacks as a performance criterion variable, including the possible influences of demand characteristics, variability in individual interpretation of the question, and intentional misreporting or deception.

Table 7. ANOVA results for Stanford Sleepiness Scale

F	df	*p*	η_p^2
26.30	(9, 126)	< .000001*	.653

* Mean difference significant at the .05 level

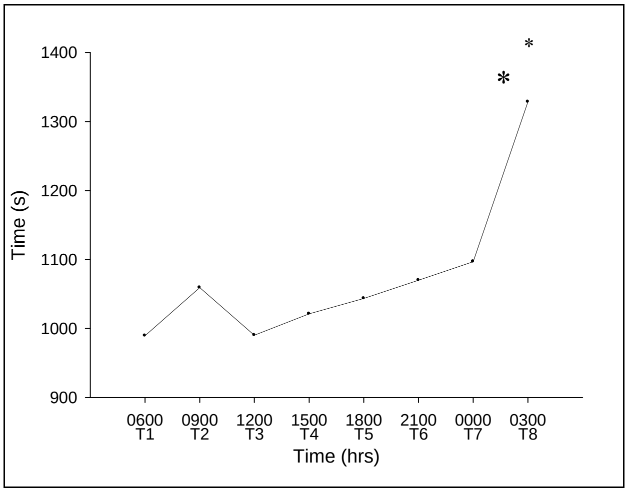

Figure 13. Mean Total Lapse Time on Flight Simulator Performance in seconds at each test trial across time. Post-hoc analyses revealed significant differences between T8 and all other trials, indicating a distinct point at which group vigilance began to fail (*). Notably, this pattern is highly similar to PVT lapses (Figure 10).

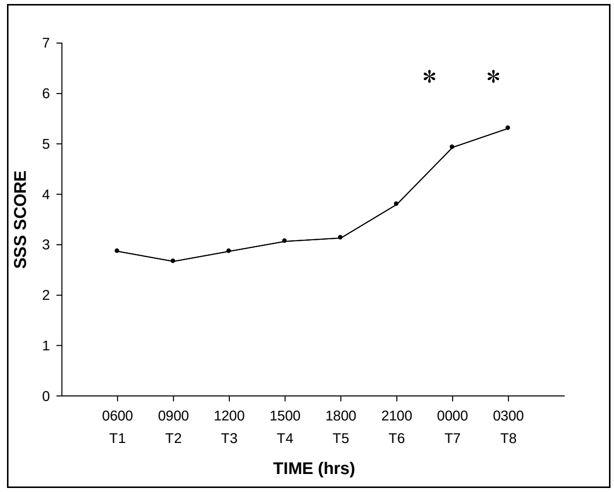

Figure 14. Mean Stanford Sleepiness Scale (SSS) score at each test trial across time. Post-hoc analyses revealed significant differences between T6, T7, and T8 and all other trials, indicating that participants felt significantly sleepier as time awake increased. The most operationally significant differences, between T6 and T7 – T8, are noted (*).

Stage 1 Summary

Stage 1 analyses were conducted in order to establish the sensitivity of each predictor and criterion variable of interest to change across time, allowing inference of the relation of group averages on those variables to time spent without sleep, and hence fatigue. These results suggest that the prospects for development of a squadron-level tool to detect aviator fatigue state (i.e. “readiness-to-fly”) in real time are good. However, the predication of Stage 1 analyses on a repeated measures design does not provide information on the causal relation between predictor and criterion variables. Group average based analyses, such as those performed in Stage 1, also mask any potential individual differences in fatigue response. We therefore constructed a series of predictive Hierarchical Linear Models (HLMs) in Stage 2, using the framework provided by Stage 1 to examine any potential individual differences in fatigue response as well as the ability of each variable to predict a criterion measure of performance, PVT lapses. Among the available performance markers, PVT lapses were deemed most appropriate for this application due its extensive validation in the fatigue literature as well as its operational relevance. For instance, vigilance of display change on the PVT response box can be likened to vigilance of display change on a radar screen. Stage 2 analyses are introduced in more detail in the next section.

Stage 2

Stage 1 longitudinal analyses revealed significant cognitive and physiological decrements attributable to fatigue on a group level. When group results were visually inspected at the individual level, two distinct clusters in the data emerged, suggesting the possibility of important individual differences in responses to fatigue. In order to statistically examine individual variability in fatigue-related performance decrements, a series of two-level Hierarchical Linear Models was performed. An exploratory bivariate analysis examined the ability of all significant Stage 1 predictors to explain PVT lapses during the test day. FAST scores were included in this analysis series, for their potential ability to predict performance on a group level as well as any possible individual differences within that ability. The results of the bivariate analyses informed subsequent moderation and multiple predictor models.

Bivariate HLMs

For all bivariate HLMs, fixed (level 1 equations) and random (level 2 equations) effects of the predictors were included, allowing determination of an overall effect of each predictor, as well as whether the relation was consistent or varied across subjects. Significance at level 1 indicates a group effect, while significance at level 2 indicates significant individual differences within that overall effect. If the random effect was not significant, indicating that there was no significant inter-individual variability, the model was refitted without the random effect of the predictor in order to focus on the group effect. Any variables that exhibited a significant bivariate relation at level 1, level 2, or both with PVT lapses are identified in Table 8. These include Time, FF_rawRT, FF_daRT, FF_shiftACC, PMI Fit 2000 Saccadic Velocity (PMI_SV), and FAST. The nature of these effects, including graphs of the individual slopes for each significant relation, is presented next.

Time. Level 1 and level 2 equations were significant for Time predicting PVT Lapses. The group effect replicated the longitudinal relation of PVT lapses across time established in Stage 1 analyses. The effect at level 2, indicating significant individual differences about the group slope, is an excellent illustration of the application of HLM to these data and the importance of considering individual fatigue responses when predicting performance. Visual inspection of Figure 15 reveals at least two distinct groups in the data when viewed as individually plotted lines. For some subjects, lapses increase at a much faster rate across time than for other subjects. Conceptualizing this difference in terms of fatigue susceptibility, individuals with high fatigue-susceptibility can be identified by their steep slopes. Low fatigue susceptible individuals exhibit the opposite trend, with little to no change in PVT lapses in relation to Time. The difference between the steepness of the two most extreme subject slopes and the rest of the group would not be evident if one line was fitted based on a group average. In this case performance decrement would be under-predicted for individuals who were actually most fatigue susceptible, and over-predicted for those who were not. Operationally this could lead to over-utilization of performance compromised individuals and under-utilization of mission-ready individuals.

Table 8. Bivariate HLMs Relations with Outcome = PVT Lapses

	Level 1				**Level 2**			
Variable	**Equation**	**t**	**df**	***p***	**Equation**	**χ^2**	**df**	***p***
Time	Y = B0 + B1*(Time) + R	2.47	14	0.03	B0 = G00 + U0 B1 = G10 + U1	50.76	14	0.00
FF_rawRT	Y = B0 + B1*(FF_RAWRT) + R	2.65	118	0.01	B0 = G00 + U0 B1 = G10 + U1	11.38	14	> 0.50
FF_daRT	Y = B0 + B1*(FF_DART) + R	2.10	118	0.04	B0 = G00 + U0 B1 = G10 + U1	13.97	14	> 0.50
FF_shiftACC	Y = B0 + B1*(FF_shift) + R	-2.76	14	0.01	B0 = G00 + U0 B1 = G10 + U1	38.88	14	0.001
PMI_SV	Y = B0 + B1*(PMI_SV) + R	-1.99	14	0.07	B0 = G00 + U0 B1 = G10 + U1	25.36	14	0.03
FAST	Y = B0 + B1*(FAST) + R	-2.82	14	0.01	B0 = G00 + U0 B1 = G10 + U1	276.37	14	0.00

Note. PVT = Psychomotor Vigilance Task, FF = Flight Fit, rawRT = Reaction Time, daRT = Divided Attention Reaction Time, shiftACC = Shifting Accuracy, PMI = Pulse Medical Instruments, SV = Saccadic Velocity, and FAST = Fatigue Avoidance Scheduling Tool.

Flight Fit Raw Reaction Time (FF_rawRT). Although visual inspection of Figure 16 may appear to suggest that there is a significant effect of the predictor at level 2, there is not enough variability among individual slopes to constitute a significant random effect. In other words, visual variability does not translate into statistically significant individual differences in this case. This underscores the importance of conceptualizing potential individual differences both visually and statistically within the context of the variable

under consideration. Because the level 2 equation was not significant using a random effect, the level 1 equation reported here consists of a re-estimation of the model without the random effect (see Table 8). The re-estimated model was significant at level 1 for FF_rawRT predicting PVT lapses, such that as reaction time increases, PVT lapses increase. The presence of a significant level 1 effect in the absence of a significant level 2 effect indicates that decline in FF_rawRT is best conceptualized on a group level, with the relation between FF_rawRT and PVT lapses tracking similarly for all subjects in this study sample.

Flight Fit Divided Attention Reaction Time (FF_daRT). As with FF_rawRT, there was no significant random effect of FF_daRT at level 2. Re-estimation of the model without the random effect produced a significant relation of FF_daRT to PVT lapses at level 1, such that as divided attention reaction time increases, PVT lapses increase (Figure 17). This pattern, similar to what was observed with FF_rawRT, indicates that no statistically significant individual differences exist in our sample in terms of performance on FF_daRT, and that all subject performance suffers similarly under fatigued conditions.

Flight Fit Shifting Accuracy (FF_shiftACC). Level 1 and level 2 equations were significant for FF_shiftACC predicting PVT lapses. The significant level 1 relation indicates that as shifting accuracy decreases, PVT lapses increase. Visual inspection of the significant inter-slope variability at level 2 shows that some individuals exhibit a much broader range of both PVT lapses and FF_shiftACC scores than others, with those exhibiting a broader range of scores displaying the greatest fatigue-related decrement as well. That is, individuals with a relatively short plot line also tend to have flat slopes, while those with longer lines tend to have steeper slopes. Practically, this means that individuals who show more variability in their performance also tend to perform worse overall. The strong clustering of plot line end points in the lower right quadrant of Figure 18 also indicates that high shifting accuracy almost always translates to a low number of PVT lapses for both high and low fatigue susceptible individuals.

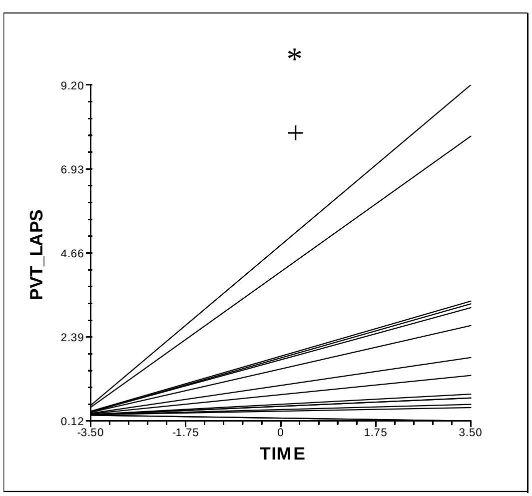

Figure 15. Individual subject slopes for PVT Lapses across trial time (group mean centered values). There was a significant group effect (*) and significant individual differences (+) such that, on average, lapses increased as time spent without sleep increased; however, the nature of that relation varied significantly from subject to subject.

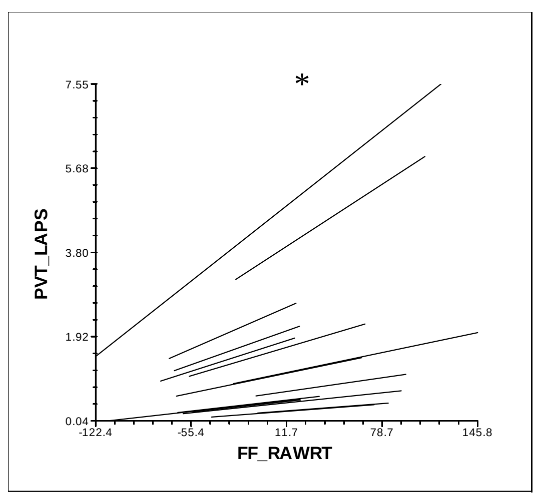

Figure 16. Individual subject slopes for PVT Lapses in relation to FF_rawRT in milliseconds (group mean centered values). There was a significant group effect (*), but no significant differences were observed among individual slopes, indicating that the relation between FF_rawRT and PVT Lapses is similar for all subjects in the sample.

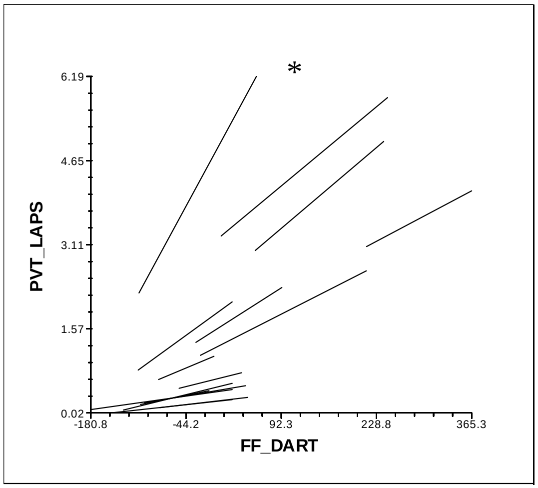

Figure 17. Individual subject slopes for PVT Lapses in relation to FF_daRT in milliseconds (group mean centered values). There was a significant group effect (*), but no significant differences were observed among individual slopes. As with FF_rawRT, this indicates that the relation between FF_daRT and PVT Lapses is similar for all subjects in the sample.

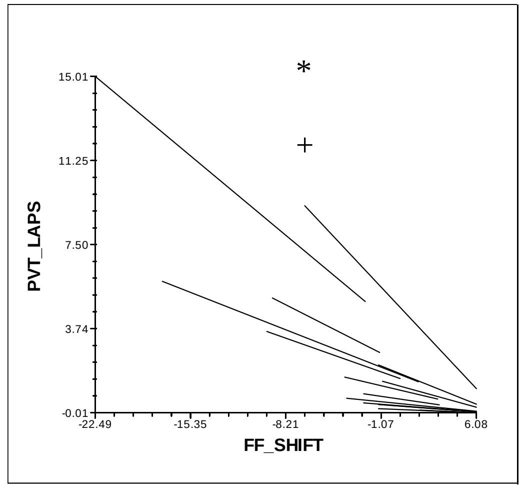

Figure 18. Individual slopes for PVT Lapses in relation to FF_shiftACC in percent (group mean centered values). There was a significant group effect (*) and significant individual differences (+) such that, on average, lapses increased as shifting accuracy decreased; however, the nature of that relation varied significantly from subject to subject.

PMI Saccadic Velocity (PMI_SV). The level 1 equation was significant for PMI_SV such that as saccadic velocity decreases, PVT lapses increase. There was also a significant random effect of the predictor at level 2. Visual inspection of the significant inter-slope variability at level 2 reveals a similar pattern to FF_shiftACC, though the dichotomy between high and low fatigue susceptibility is not as clear. Individuals who show more variability in their performance tend to perform worse overall, though this trend is not as strongly tied to low scores in the predictor variable as it is with FF_shiftACC. Some individuals with relatively slow saccadic velocity commit few lapses. The presence of these individuals, who contrast the general relation of slow saccadic velocity equaling more lapses, highlights the dynamic role of performance baseline and individual variation in saccadic velocity in relation to fatigue progression. In terms of baseline performance, individuals who start with few to no lapses tend to stay that way; graphically these are the plot lines with low intercepts. Individuals with higher intercepts, and therefore poorer baseline performance, tend to get worse across trials. The possible role of baseline performance on a measure as a predictor of fatigue-related decline across time will be discussed in more detail in the section *Moderation HLMs using Baseline PVT Performance.* In terms of individual variation, the fact that slow saccadic velocity can be, but isn't always, associated with a high number of lapses further emphasizes the need for establishing individual baselines in physiological fatigue measures (Figure 19).

FAST. The level 1 equation was significant for FAST, indicating that as FAST predicts a drop in performance, a drop in PVT vigilance occurs. There was also a significant random effect of the predictor at level 2. As with Time, this significant inter-slope variability at level 2 reveals some distinct groups: 1) those for which fatigue related decrement is well predicted, 2) those for which it is over predicted, and 3) those for which it is under predicted. Group 1 can be seen in the plot lines grouped around the center of Figure 20, where an incremental change in FAST relates to a relatively equal incremental change in PVT lapses. Group 2 is represented by the flat lines across the bottom of Figure 20, where performance decrements predicted by FAST do not materialize. Group 3 is the most striking, represented by the steep sloped lines running distinctly separate from the other plot lines. Here, actual performance suffers at a much greater rate than what is predicted by FAST. Operationally, the only acceptable predictive ability is for individuals in Group 1. As previously noted, over-prediction can result in inefficient use of manpower, and under-prediction can create a hazardous working environment.

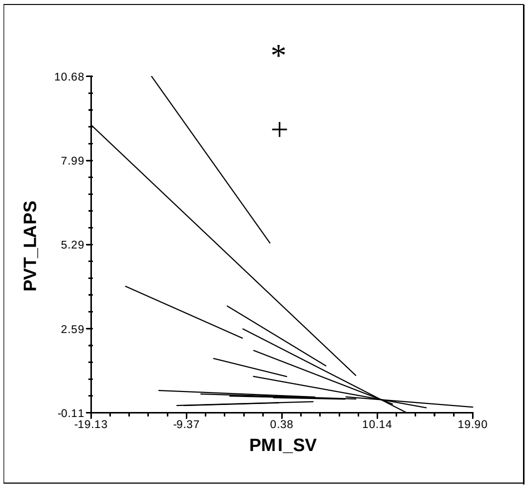

Figure 19. Individual slopes for PVT Lapses in relation PMI_SV in millimeters per second (group mean centered values). There was a significant group effect (*) and significant individual differences (+) such that, on average, lapses increased as saccadic velocity decreased; however, the nature of that relation varied significantly from subject to subject.

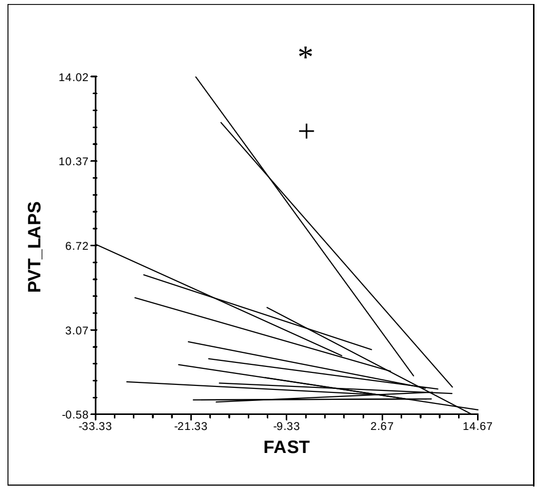

Figure 20. Individual slopes for PVT Lapses in relation to predicted performance in FAST (group mean centered values). There was a significant group effect (*) and significant individual differences (+) such that, on average, a predicted drop in performance by FAST translated into an increase in PVT Lapses; however, the nature of that relation varied significantly from subject to subject.

The Relation between the SSS and PVT Lapses

The final bivariate relation examined was between two conceptual outcome variables, PVT lapses and the SSS. While not defined as outcome and predictor *a priori*, this relation is theoretically interesting in that it allows determination of whether an individual's subjective evaluation of their fatigue state from the SSS is predictive of their objective fatigue-related performance on the PVT. Level 1 and level 2 equations were significant for the SSS predicting PVT lapses, indicating that, in general, subjective sleepiness is predictive of PVT vigilance, but there are significant individual differences in that general relation (see Figure 21). Visual inspection of Figure 21 reveals that all subjects report getting progressively more tired. However, for the individuals represented by lines with flat slopes, increasing sleepiness does not correlate with an actual drop in performance. For subjects with relatively steep slopes subjective sleepiness is related to performance decrements. Again, we are presented with low and high fatigue susceptible subjects. The operational impact of this relation is the clearest of the Stage 2 analyses: asking someone how sleepy they are holds variable diagnostic value in terms of predicting subsequent performance. Operationally this emphasizes the importance of objective fatigue measurement, such as with the metrics currently evaluated in this report.

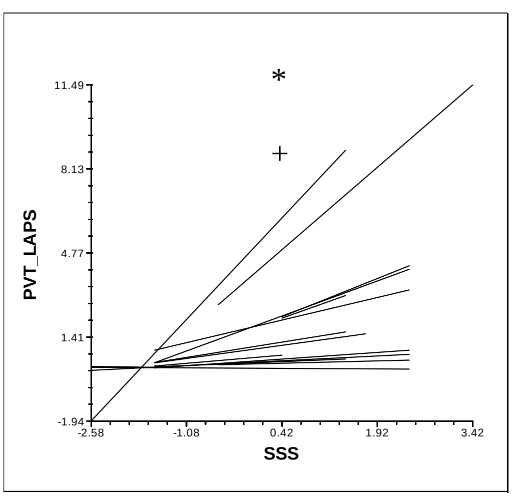

Figure 21. Individual slopes for PVT Lapses in relation to SSS scores (group mean centered values). There was a significant group effect (*) and significant individual differences (+) such that, on average, as subjective sleepiness increased, PVT Lapses increased; however, the nature of that relation varied significantly from subject to subject.

Moderation HLMs using Baseline PVT Performance

Examining the bivariate relations as a group, the largest effects of fatigue, as well as the largest relations of predictors explaining fatigue, appear to be among subjects who had the largest number of lapses. This trend suggests that variability in the relations may be explained by the participants' average baseline PVT performance from the Practice phase. Using the bivariate analyses that had a significant random effect of the predictor, a series of analyses in which average number of PVT lapses during the Practice phase was included as a level 2 variable. A significant effect of Practice phase PVT lapses at level 2 would indicate a moderating influence of baseline performance on subsequent performance when predicted by the level 1 variable. For example, prediction of PVT lapses by FF_shiftACC while fatigued can be partially explained by an individual's baseline PVT performance (see Table 9), such that a higher number of baseline lapses is significantly predictive of a steeper slope representing the relation between FF_shiftACC and PVT lapses while fatigued. The same pattern is true when baseline PVT lapses are used as a moderator between FAST and PVT lapses while fatigued (see Table 9). Practically, this means that the worse a person's vigilance is when rested, the more extreme their negative reaction to fatigue will be when subjected to sleep loss. Operationally, this means that a warfighter's vigilance while fatigued can be predicted, at least in part, by their vigilance while rested, above and beyond the considerable predictive ability of FF_shiftACC and FAST.

Table 9. Moderating Effects of Baseline Lapses on Performance Outcome = PVT Lapses

	Level 1				**Level 2**			
Variable	**Equation**	**t**	**df**	***p***	**Equation**	**χ^2**	**df**	***p***
FF_shiftACC	Y = P0 + P1*(FF_shiftACC) + E	-2.06	116	0.04	P0 = B00 + B01 *(BLAPSAVG) + R0 P1 = B10 + B11 *(BLAPSAVG) + R1	29.01	13	0.01
FAST	Y = B0 + B1*(FAST) + R	-3.876	116	0.00	B0 = G00 + G01 *(BLAPSAVG) + U0 B1 = G10 + G11 *(BLAPSAVG) + U1	230.14	13	0.00

Note. PVT = Psychomotor Vigilance Task, FF = Flight Fit, RT = Reaction Time, shiftACC = Shifting Accuracy, and FAST = Fatigue Avoidance Scheduling Tool.

Multivariate HLM

Bivariate analyses established the significant relations of six individual predictors and one outcome variable to PVT lapses. Many of the significant predictors are conceptually related, such as FF_shiftACC and PMI_SV, and may share statistical explanatory variance. A multivariate HLM using all significant bivariate predictors except Time was therefore constructed. Time was excluded as it is assumed to be theoretically and statistically collinear with the other predictors. PVT lapses were used as the outcome. The purpose of this analysis was to determine which, if any, variables possessed unique predictive ability above and beyond the others. Results indicate that out of FF_rawRT, FF_daRT FF,_shiftACC, PMI_SV and FAST, FF_daRT, FF_shiftACC, and FAST remained significant predictors of PVT lapses at level 1, while FF_rawRT and PMI_SV dropped out. This suggests that the significant predictive ability of FF_rawRT and PMI_SV may already be captured by aspects of FF_daRT, FF_shiftACC, and FAST.

Stage 2 Summary

Stage 2 analyses were conducted to establish the ability of significant Stage 1 variables to predict performance on the PVT. HLM was used in order to observe significant relations at the group and individual levels simultaneously. Three aspects of Flight Fit (FF_rawRT, FF_daRT, and FF_shiftACC), one of PMI Fit 2000 (PMI_SV), and FAST were able to significantly predict PVT lapses at the group level. Of these,

FF_shiftACC, PMI_SV, and FAST also displayed significant individual differences in their relation to PVT lapses. There was also significant inter-individual variability in the relation between SSS scores and PVT lapses, uncovering a disconnect between subjective self-report of fatigue and its objective consequences. Moderation analyses revealed that baseline PVT lapses can be used to predict the relation between shifting accuracy and PVT lapses, and a comprehensive multivariate HLM demonstrated that divided attention reaction time, shifting accuracy, and FAST predict significant systematic variance in PVT lapses above and beyond raw reaction time and saccadic velocity. These results clarify two important points for the development of an individualized readiness-to-fly fatigue measure. First, fatigue measurement *must* take individual differences into account. This cannot be accomplished using group norms as the comparison baseline for individual prediction. Though a group-based approach will give successful approximations for most people, the statistical outliers are actually the most critical to capture when trying to predict fatigue-related performance in an operational context. The most efficient way to capture these outliers would be to establish individual baselines of performance and then track changes from that baseline much in the same way the PMI Fit 2000 does, with focus on the significant predictors found in this stage. The moderation results suggest that an individual's rate of decline due to fatigue could be predicted while establishing baseline vigilance ability. Second, the results of the multivariate HLM demonstrate the need to carefully balance predictive power and practical application. While raw reaction time and saccadic velocity do not explain systematic variance in PVT lapses above and beyond the other significant predictors, they are the fastest assessments and least obtrusive of the performed tests. To further inform the balance between predictive power and practical application, Stage 3 consisted of an incremental validity analysis to evaluate the predictive ability of different conceptual combinations of variables with respect to operational utility.

Stage 3

The multivariate HLM from Stage 2, including group and individual difference effects, is difficult to translate into a single fatigue prediction algorithm. Ideally, accurate prediction would be based on an individual equation for each subject in which the respective beta weights for each variable change according to inter-individual slope variability. The individualized algorithm approach is beyond the scope of this report, though the significant level 1 equations from the Stage 2 multivariate HLM are further examined here. Using level 1 results from Stage 2 analyses, a group-based scoring algorithm was constructed. It is important to note that this approach has the same strengths and weaknesses as other currently used group- based prediction algorithms. However, its conceptual use for this report is not necessarily in creating a mission-ready fatigue prediction algorithm; rather, it is in examining the interaction and respective contribution of cognitive and physiological measures of fatigue in an incremental fashion in a single equation. To further explore the incremental validity of each significant Stage 2 predictor at the group level, a series of enter-method General Linear Models (GLMs) was constructed. Because an established group-based scoring algorithm is already included in this report (FAST), the first analysis compared variance in PVT lapses explained by FAST alone with total variance explained when significant Flight Fit and PMI predictors were included with FAST. Results are presented in Table 10. As in the Stage 2 analysis, FAST was able to explain a significant amount of variance in PVT lapses, about 14% (R-square = .138), at the group level. Addition of raw reaction time, divided attention reaction time, shifting accuracy, and saccadic velocity increased the amount of variance explained to about 36% (R-square = .357), a significant change statistically and conceptually.

Table 10. Incremental Ability of FAST, Flight Fit Subscores, and PMI FIT 2000 Subscores to Predict Variance in PVT Lapses

Model	Variables	Equation	R Square	Δ *F*	*df1*	*df2*	*p*
1	FAST	PVT_Lapses = FAST * -.371	.138	18.863	1	118	.000
2	FAST FF_rawRT FF_daRT FF_shiftACC PMI_SV	PVT_Lapses = (FAST * -.126) + (FF_rawRT RT * .029) + (FF_daRT * .03) + (FF_shiftACC * -.424) + (PMI_SV * - .211)	.357	9.688	4	114	.000

Note. FAST = Fatigue Avoidance Scheduling Tool; FF_rawRT = Flight Fit raw reaction time; FF_daRT = Flight Fit Divided Attention Reaction Time; FF_shiftACC = Flight Fit shifting accuracy ; PMI_SV = PMI Saccadic Velocity. All equation values assumed to be group mean centered.

GENERAL DISCUSSION

The results of the current investigation underscore four main points: fatigue measurement and prediction must take individual differences into account, optimal fatigue measurement requires considering objective cognitive *and* physiological aspects, operational utility is promising for aspects of both Flight Fit and PMI FIT 2000, and additional research is needed to establish potential for implementation of these test instruments.

Individual Differences in Measuring and Predicting Fatigue

The results strongly suggest that individual differences in fatigue susceptibility must be taken into account when measuring and predicting fatigue. Individual differences are central to conceptualizing fatigue in operational contexts where understanding a service member’s strengths and weaknesses is key to optimizing task assignment and safety. Other fatigue researchers have posited that fatigue susceptibility is a trait-like characteristic (Caldwell, 2005), which has systematic, identifiable neurobiological and physical underpinnings (Caldwell et al., 2005; Rétey et al., 2006; Killgore et al., 2009) that may be modified through training (Klingberg, Forssberg, & Westerberg, 2002). Though performance prediction based on a group average accounts for most individuals, those who are not properly categorized under such an approach are theoretically and practically the most critical to capture. For instance, those who are highly susceptible to fatigue may require additional training, or more tailored scheduling, similar to physical readiness training. Those who are highly resistant to fatigue may be better suited to situations in which sustained vigilance is routinely required, such as in Air Traffic Control (ATC) – much like assigning certain duties according to physical strength.

Practically, these results suggest that quantifying individual differences can be easily accomplished by establishing individualized baselines of performance, as illustrated by the PMI Fit 2000. By taking 10 rested baseline readings prior to sleep deprivation, the PMI system was able to calculate each subject’s performance in terms of a departure from that subject’s own record, no matter where that record began and no matter the rate at which that departure proceeded.

Measuring and Predicting Fatigue with Objective Cognitive and Physiological Aspects

Cognitive (e.g., tests of vigilance, working memory) and physiological (e.g., fMRI, EEG) measurements have been used previously to track fatigue-related changes, and previous results verify that they are similarly affected by sleep deprivation (Berka et al., 2007). Biomathematical models of sleep deprivation and performance emphasize that cognitive and physiological aspects of fatigue are numerous, interdependent, and complex (Dinges, 2004; Rétey et

al., 2006). The results from the current study confirm that the predictive ability of an already biomathematically-based model, FAST, was significantly improved by adding additional individualized cognitive (Flight Fit) and physiological (PMI FIT 2000) measures. Operationally, the measures tested were fast, effective, and adaptable to fatigue vulnerability across and within individuals. Future individualized fatigue detection tools should incorporate individualized cognitive and physiological measurements to maximize predictive ability and successful categorization.

Results also suggest that these individualized cognitive and physiological measurements should be as objective as possible. The design included the SSS as a subjective, self-report measure of fatigue to observe the relation between an individual's feelings of fatigue and their actual performance while fatigued. While it doesn't include a self-rating of performance or performance potential *per se*, and therefore excludes an analysis of an individual's ability to perceive or predict their actual performance, it does quantify an individual's perception of their general fatigue state. Results reveal a closely clustered group average for SSS across time, such that all subjects reported getting more tired as time awake increased. However, the individual differences in performance remain, meaning that fatigue resistant individuals still get sleepy - they just continue to perform at baseline levels despite their sleepiness. These results suggest that, simply put, asking an individual whether they are too tired to perform has little diagnostic value for actual performance, especially for the individuals who would most likely do well. This point further emphasizes the need for multi-dimensional, objective evaluation and prediction of performance while fatigued.

Usability and Recommendations

Aspects of Flight Fit and PMI FIT 2000 are promising for use as valid real-time readiness-to-fly assessment tools in Naval Aviation squadrons, but key adjustments need to be made to the manufacturers' current scoring algorithms. For both instruments, the manufacturers' current scoring algorithms are inadequate for fatigue detection in a Naval Aviator population. In order to detect significant results, analyses had to be performed using raw scores from both tests.

Flight Fit. Initial analyses for Flight Fit, presented in an interim report to the sponsor, were based on the program's standardized output. This output was presented to subjects at the end of each testing session as a percentile rank of performance in relation to established norms (i.e., 90th percentile for shifting accuracy, etc.) Unfortunately, the Flight Fit test battery was normed by the manufacturer using a sample of truck drivers, not Naval Aviators. Observed percentile rank scores in the present Naval Aviator sample clustered near the high end of truck-driver derived normative scores. The resulting restriction in range of observed normed scores resulted in no measurable fatigue effects. When raw scores were used, as in the analyses presented in this report, highly significant fatigue effects were exposed on multiple component subtests of Flight Fit. Before Flight Fit or any test battery can be utilized as a readiness-to-fly screener for Naval Aviators, new scoring norms based on Naval Aviator performance for significant component subtests must be established.

PMI FIT 2000. For PMI FIT 2000, the failure to predict fatigue using the manufacturer's scoring algorithm may be due to deviation from the instrument's primary intended use: detection of impairment due to drugs and alcohol. Analysis of the raw data revealed saccadic velocity to be especially sensitive to fatigue effects. However, the instrument's other pupillometric indices and overall "FIT Index", which are known to be sensitive to the effects of drugs and alcohol, were unrelated to fatigue in our sample. The manufacturer has developed alternate setting since the start of this study called Fatigue Analyzer mode that is based heavily on saccadic velocity. Future studies are planned using the Fatigue Analyzer mode to further validate the use of the PMI Fit 2000 as a readiness-to-fly screener in Naval Aviation.

Though NAMRL cannot recommend use of Flight Fit and the PMI FIT 2000 instruments in their current form, additional testing designed to re-norm Flight Fit with a Naval Aviation population using only the significant component subtests reported here, and further testing with the PMI FIT 2000 in Fatigue Analyzer mode, are both highly recommended.

Next Steps / Future Directions

The current results suggest that aspects of Flight Fit and PMI FIT 2000 warrant further investigation in order to better determine their usefulness to the Fleet as individualized, readiness-to-fly screeners. Beyond the need to adjust or modify the scoring algorithms currently provided by the manufacturers, there are additional questions to examine in regard to the operational utility of these tools. For instance, sleep deprivation is not always encountered in a sustained, acute manner. Gradual, chronic sleep restriction, in which a service member may only get a few hours of sleep a night during intensive training or in high-tempo operations, must also be considered.

Summary

Over the course of 25 hours of continual wakefulness in a laboratory setting, occulometric measures of saccadic velocity and cognitive performance on a multi-faceted test battery were significantly sensitive to fatigue. More importantly, these tools were able to identify an individual’s susceptibility to performance decrements associated with fatigue, a capability not available with tools based on average group performance. While these results are promising, further evaluation across a wider array of individuals, settings, and fatigue durations is needed prior to military implementation. The ultimate goal is a comprehensive and field-expedient tool for transition to the Fleet, capable of providing an accurate assessment of specific fatigue states at the level of the individual warfighter. This would reduce the negative impact of fatigue on performance and inform a commander’s decision-making on manning and mission readiness.

Acknowledgements: The authors would like to express their sincere gratitude to Dr. Jamie DeCoster, CDR Rita G. Simmons, and Andrew Reinert for their technical advice and assistance during the preparation of this report.

References

Balkin, T.J., Bliese, P.D., Belenky, G., Sing, H., Thorne, D.R., Thomas, M., Redmond, D.P., Russo, M. & Wesensten, N.J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research, 13*, 219-227.

Belland, K. M & Bissell, C. (1994). A subjective study of fatigue during navy flight operations over southern Iraq: Operation Southern Watch. *Aviation, Space, and Environmental Medicine, 65*, 557-561.

Berka, C. , Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T. et al. (2007). EEG coorelates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine, 75 (5 Suppl)*, B231-44.

Caldwell, J. A. (2004). Fatigue in aviation. *Travel Medicine and Infectious Disease, 3*, 85-96.

Caldwell, J. A., Mu, Q., Smith, J. K., Mishory, A., Caldwell, J. L., Peters, G., et al. (2005). Are individual differences in fatigue vulnerability related to baseline differences in cortical activity? *Behavioral Neuroscience, 119*, 694-707.

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., Neri, D. F. (2009). Fatigue countermeasures in aviation, *Aviation, Space, and Environmental Medicine, 80*, 29-59.

Dinges, D. F. (2004). Critical research issues in development of biomathematical models of fatigue. *Aviation, Space, and Environmental Medicine, 75 (3 Suppl)*, A181-91.

Dinges, D. F., Mallis, M. M., Maislin, G., Powell, J. W. (1998). *Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management.* Final report for the U.S. Department of Transportation, National Highway Traffic Safety Administration. NHTSA Report No. HS 808 762.

Graybiel, A., Brown, N., & Crispell, R. (1943). *A statistical analysis of the medical complaints of flying personnel at the naval air station, Pensacola, Fla.* DTIC Report AD0620281.

Graybiel, A., Horwitz, O., & Gates, D. (1944). The problem of fatigue among student pilots at the naval air training center, Pensacola, Florida, *Aviation Medicine*, 9-25.

Hoddes, E., Dement, W., & Zarcone, V. (1972). The development and use of the Stanford sleepiness scale. *Psychophysiology, 9*, 150.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., et al., (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine, 75*, A44-53).

Killgore, W.D.S., Grugle, N.L., Reichardt, R.M., Killgore, D.B., & Balkin, T.J. (2009). Executive functions and the ability to sustain vigilance during sleep loss. *Aviation, Space, and Environmental Medicine, 80(2)*, 81-87.

Lyons, T. J. & French, J. (1991). Modafinil: the unique properties of a new stimulant. *Aviation, Space, and Environmental Medicine, 62*, 432-5.

Myrick, H., Malcom, R., Taylor, B., & LaRow, S. (2004). Modafinil: Preclinical, clinical, and post-marketing surveillance - a review of abuse liability issues. *Annals of Clinical Psychiatry, 16* (2), 101-109.

Naval Safety Center (2006). Aeromedical factors involved in Naval aviation class A flight mishap from FY00-06. Unpublished briefing slide. Norfolk, VA.

Phillips, J.B., Arnold, R.D., Strompolis, M., & Simmons, R.G. (2009). The efficacy of armodafinil for sustained performance among Navy air traffic controllers eight to twelve hours post dose. NAMRL Technical Report 09-18, Naval Aerospace Medical Research Laboratory, Pensacola, FL.

Rétey, J. V., Martin, A., Gottselig, J. M., Khatami, R., Durr, R., Achermann, P., et al. (2006). Adenosinergic mechanisms contribute to individual differences in sleep deprivation-induced changes in neurobehavioral function and brain rhythmic activity. *The Journal of Neuroscience, 26*, 10472-9.

Russo, M., Thomas, M., Sing, H., Thorne, D. Balkin, T., Wesensten, N., et al, (1999). Saccadic velocity and pupil constriction latency changes in partial sleep deprivation, and correlations with simulated motor vehicle crashes. *Sleep, 22 (Suppl. 1)*: S297-298.

Van Dongen, H. P., Baynard, M. D., Maislin, G. & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep, 27*, 423-33.

Van Dongen, H. P., Caldwell, J. A., & Caldwell, J. L. (2006). Investigating systematic individual differences in sleep-deprived performance on a high-fidelity flight simulator. *Behavior Research Methods, 38*, 333-43.

Van Dongen, H. P., Rogers, N. L., & Dinges, D. F. (2003). Sleep debt: Theoretical and empirical issues. *Sleep and Biological Rhythms*, *1*, 5-13.

APPENDIX A. List of Abbreviations

FF_rawRT- Flight Fit Reaction Time

FF_vsRT- Flight Fit Visual Scanning Reaction Time

FF_vsACC- Flight Fit Visual Scanning Accuracy

FF_daRT- Flight Fit Divided Attention Reaction Time

FF_daACC- Flight Fit Divided Attention Accuracy

FF_SRT– Flight Fit Shifting Reaction Time

FF_shiftACC- Flight Fit Shifting Accuracy

FF_fdRT- Flight Fit Focus Reaction Time in the Presence of Distracters

FF_STM- Flight Fit Short Term Memory

PMI_SV - PMI FIT 2000 Saccadic Velocity

APPENDIX B. Confidential Medical Questionnaire

Screening Number:____________ Participant Number: ___(for office use only) Date: _____________

Gender (check one): Male ☐ Female ☐ Age: _____________ Height: ______________ Weight: ____________

<u>Directions</u>: **Circle "Yes"** or **"No"**. ***These questions are being asked to ensure your safety in this study.***

Do you have a current flight physical?	Yes	No
Have you ever been diagnosed with any significant medical problems? (e.g., heart/circulatory disease)	Yes	No
Have you ever been diagnosed with any neurological syndrome, disorder, or injury? (e.g., migraines, epilepsy, or traumatic brain injury)	Yes	No
Have you ever been diagnosed with any psychiatric disorder? (e.g., depression or anxiety)	Yes	No
Have you ever been diagnosed with any sleep related disorders (e.g., sleep apnea, insomnia, narcolepsy, sleep walking)	Yes	No
Have you used any tobacco products in the last 6 months? If yes, please list quantity, frequency and type of product: _____________________________	Yes	No
Do you take any prescribed medication on a regular basis? If yes, please list : ____	Yes	No
Have you consumed any caffeine within the past 48 hours? If yes, how much? ____ Is this your normal amount? ___________________________________	Yes	No
Have you consumed any alcohol within the past 48 hours?	Yes	No
Indicate all medication you have used in the past 24 hours. *(circle all that apply)* a. None d. Antihistamines b. Sedatives/Tranquilizers e. Decongestants c. Aspirin/Tylenol/any analgesic f. Other (please specify) ___________________________		
How many hours did you sleep last night? ___________ Was this amount sufficient?	Yes	No
Females: Are you currently pregnant or lactating?	Yes	No

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YY)
25-05-10

2. REPORT TYPE
Technical Report

3. DATES COVERED (from – to)
Jan 2009-Jan 2010

4. TITLE
Preliminary Validation of a Readiness-to-fly Assessment Tool for Use in Naval Aviation

5a. Contract Number:
5b. Grant Number:
5c. Program Element Number:
5d. Project Number:
5e. Task Number:
5f. Work Unit Number: 70803

6. AUTHORS
Chandler, Joseph F., Ph.D, Arnold, Richard D., Ph.D, Phillips, Jeffrey B., Ph.D
Lojewski, Renee A., M.A., Horning, Dain S., B.S.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
280 Fred Bauer Street Building 1811
Pensacola, FL 32508

8. PERFORMING ORGANIZATION REPORT NUMBER
NAMRL 10-22

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Bureau of Medicine and Surgery
Department of the Navy
2300 E Street, NW
Washington, DC 20372-5300

10. SPONSOR/MONITOR'S ACRONYM(S)
BUMED

11. SPONSOR/MONITOR'S REPORT NUMBER(s)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Fatigue is the most frequently cited physiological factor contributing to the occurrence of US Naval Aviation Class A flight mishaps. Accordingly, the Naval Safety Center (NSC) has identified the need for a quickly-administered individualized fatigue assessment tool to determine a pilot or aircrew member's readiness to fly. The Naval Aerospace Medical Research Laboratory conducted validation research on Flight Fit – a series of computer administered cognitive tasks sensitive to fatigue, and PMI FIT 2000 – a physiological test of oculometric properties linked to fatigue, for their potential to serve as individualized fatigue detection tools. Performance on both assessments was observed in concordance with performance on a suite of industry standard fatigue-sensitive measures (e.g., the Psychomotor Vigilance Test) at regular intervals over 25 hours of continual wakefulness in naval aviators. Results indicate significant group and individual differences related to fatigue for several aspects of both measures, and suggest that with appropriate adjustments, both Flight Fit and PMI FIT 2000 could serve as valid real-time readiness-to-fly assessment tools in Naval Aviation squadrons. Follow-on studies to determine the exact nature of these adjustments and usability of the tools in their current form are discussed.

15. SUBJECT TERMS
Fatigue, physiological readiness, oculometrics, cognitive fatigue assessment, psychomotor vigilance task

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCL	UNCL	UNCL	UNCL	27

18a. NAME OF RESPONSIBLE PERSON
Officer in Charge

18b. TELEPHONE NUMBER (INCLUDING AREA CODE)
Commercial: 850-452-3486, DSN: 922-3486